

Sistemas de ficheros

Orientación a los Procesos

Surge de la automatización los procesos manuales que se realizaban en cada departamento de las organizaciones.

- Los datos de cada archivo están fuertemente vinculados al programa que los procesa.
- Los procesos y los datos se almacenan en cintas magnéticas. Para el procesamiento, los procesos y los datos se cargan en el equipo. Después del procesamiento, los datos resultantes se vuelcan de nuevo en la cinta.
- Cada proceso maneja sus propios archivos:
 - Ficheros con datos de entrada, con los datos a procesar.
 - Ficheros con datos de salida, con los datos resultantes de realizar el procesamiento.
- Cada archivo contiene datos relacionados con un proceso determinado.
- Los datos de cada fichero son independientes de otros archivos y de otros procesos.

Cada archivo de datos está atado a un programa específico. Los datos se guardan en cintas magnéticas, se cargan para su procesamiento y luego se vuelcan de nuevo. Cada proceso maneja sus propios archivos: entrada y salida. Cada archivo está relacionado con un proceso y es independiente de los demás.

Registros

En los ficheros, los datos se organizan en forma de registros.

- Cada registro agrupa un conjunto de datos relacionados relativos al mismo objeto o elemento.
- Cada dato se corresponde con una categoría que le da significado.

Ejemplo:

- Un registro contiene los datos de un cliente: los datos de Javier Barquín.
- Cada dato corresponde a un campo o atributo: el DNI, el nombre, el teléfono, ...

Registros de tamaño fijo:

- Todos los ficheros tienen el mismo tamaño.
 - Todos los ficheros tienen los mismos campos.
 - Cada campo tiene un espacio reservado en memoria.
- El valor de un campo no puede exceder el tamaño reservado.
- Si un campo no tiene valor o tiene un valor pequeño, ocupará igualmente el tamaño máximo, lo que conlleva un desperdicio de espacio de almacenamiento.
- Es fácil calcular la posición de un fichero n : bastará sumar $n \times \text{tamaño}$ a la posición inicial.

Registros de tamaño variable:

- Os ficheros pueden tener tamaños distintos.
- El tamaño de un campo será el que precise su valor.
- Se ahorra espacio de almacenamiento.
- No podemos determinar dónde empieza y termina cada fichero: es necesario usar algún sistema para identificarlo (por ejemplo, marcas de fin de fichero).
- Conocer la posición en memoria de cada fichero es complejo.
- Cambios en un fichero exigen reasignar memoria.

Estructura de los ficheros

Existen distintas formas de representar los ficheros, la más habitual es en forma de tabla:

- Cada fila representa un fichero.
- Cada columna representa un campo.

La otra es la que parecen fichitas tipo UML.

Inconvenientes de los sistemas de ficheros

- **Redundancia:** Los datos que se utilizan en varios procesos estarán repetidos en varios ficheros.

Por ejemplo, tener dos listas de la compra iguales.

- **Costo de almacenamiento elevado:** Cada dato se almacena varias veces, cada copia ocupa espacio en el disco.

Por ejemplo, tener tantas listas de la compra que no te caben las llaves

en el bolsillo.

- **Tiempos de procesamiento elevados:** El tratamiento de los datos es generalmente secuencial. El procesamiento de un solo dato implica recorrer ficheros completos.

Por ejemplo, tardas media hora en encontrar la lista de la compra porque tienes que leer las todas las listas de la compra que hicistes desde enero.

- **Modificación costosa:** Si un dato cambia, debe modificarse en todos los ficheros en los que esté.

Por ejemplo, tienes que modificar "piña" donde pone "pizza" y tienes que hacerlo en las 189 listas de la compra que tienes.

- **Inconsistencias:** Se produce una inconsistencia si un dato se repite en varios ficheros con distinto valor. Generalmente, esto se debe a que un cambio no se reprodujo en todas las copias del dato.

Por ejemplo, en una lista de la compra pone "Leche desnatada", y en otra "Leche entera"; cuando llegues al supermercado no sabes cuál comprar.

Tipos de ficheros:

Tipos según su uso

- **Mestres:** Son los principales sistemas de información y contienen los datos fundamentales. Raramente cambian.

Ejemplo: Un sistema de gestión de clientes en una empresa. Los datos de los clientes, como nombres, direcciones y números de teléfono, se almacenan en un archivo maestro. Estos datos son fundamentales y cambian raramente, ya que no es común que los clientes cambien su información personal con frecuencia.

- **Constantes:** Contienen información que apenas varía a lo largo del tiempo.

Ejemplo: Una base de datos de productos en una tienda en línea. La información sobre los productos, como sus nombres, descripciones y precios, se almacena en un archivo constante. Estos datos apenas varían con el tiempo, ya que los productos no cambian sus características esenciales con frecuencia.

- Históricos: Almacenan copias de los ficheros mestres que se realizan periódicamente.

Ejemplo: Un sistema de copias de seguridad en una empresa. Se realizan copias periódicas de los datos críticos del negocio, como registros financieros, y se almacenan en archivos históricos con marcas de tiempo. Estos archivos sirven como una copia de seguridad de los datos en caso de pérdida o daño y se actualizan regularmente.

- De movimiento: Almacenan los cambios que se harán en los ficheros mestres. Se eliminan después de realizar esos cambios.

Ejemplo: Un sistema de seguimiento de pedidos en una tienda en línea. Cuando un cliente realiza un pedido, se registra inicialmente en un archivo de movimiento. Este archivo contiene detalles del pedido actual, como los productos solicitados y la información de envío. Una vez que el pedido se ha procesado y entregado, se elimina del archivo de movimiento, ya que ya no es relevante.

- De maniobra: Archivos auxiliares empleados por el software que gestiona los datos. Se destruyen tan pronto como finalizan las aplicaciones.

Ejemplo: Un sistema de almacenamiento en caché en un navegador web. Los archivos temporales, como imágenes y archivos CSS, se almacenan en un archivo de maniobra para acelerar la carga de páginas web. Estos archivos son auxiliares y se eliminan tan pronto como se cierra el navegador o se borran manualmente para liberar espacio en disco.

Tipos de fichero según la forma de acceso a los registros

Métodos de acceso a ficheros:

- Secuencial: Los registros están ordenados secuencialmente en la misma secuencia en la que se registraron. Para acceder a un registro es necesario recorrer todos los registros anteriores.
- Secuencial encadenado: Son ficheros secuenciales gestionados usando punteros. Un puntero es un dato especial que contiene la dirección de otro registro en el fichero.
- Directo o aleatorio: Permiten acceder directamente a cualquier registro del fichero, pero es necesario conocer la posición en la que se encuentra. La posición puede indicarse en bytes o mediante una clave.
- Indexado: Los registros del fichero están agrupados en bloques de registros ordenados. Para acceder a los registros se utiliza un índice en forma de tabla, que relaciona la primera clave de cada bloque con la posición en la que se encuentra. Una vez localizado el bloque, se recorren secuencialmente sus registros hasta encontrar el buscado.
- Dinámico: Permite el acceso directo o por índice a algún registro, y a partir de ese se accede a los demás de forma secuencial.

Fichero Secuencial

Características de escritura de registros:

- Los registros se escriben en el dispositivo de almacenamiento en posiciones físicamente contiguas, sin dejar espacios entre ellos, en la misma secuencia en la que se introdujeron.
- Este tipo de ficheros pueden emplear dispositivos o soportes de acceso secuencial o no direccionables, como cintas de almacenamiento magnético de datos.

Operaciones:

- Recorrido del fichero: Los registros se recorren secuencialmente empezando por el principio del fichero y hasta llegar al final.
- Acceso a un registro: Requiere crear un proceso en el que se recorre el fichero y se compara el valor del campo a localizar con el valor del mismo campo de los registros recorridos.
- Consulta de un registro: Requiere primero acceder al registro. Después la lectura es directa.
- Inserción dun registro: Solo puede hacerse al final del fichero. No se pueden insertar nuevos registros entre los ya existentes.
- Borrado dun registro: Requiere primero acceder al registro. Después el registro no se puede eliminar físicamente, sino que se marcará como borrado (borrado lógico). Los registros con la marca de borrados no se tienen en cuenta en las otras operaciones.
- Modificación de un registro: No es posible la modificación. En su lugar se procederá a borrar el registro e insertarlo de nuevo, con el campo modificado.

- Para mantener ordenado y compacto el fichero, es necesario realizar periódicamente operaciones de mantenimiento: copiar los registros no borrados en otra posición de memoria.

Ventajas:

- Acceso rápido a un bloque de registros contiguos.
- No se desperdicia espacio de almacenamiento en el dispositivo, porque no hay huecos.
- Se puede usar cualquier tipo de registro: longitud fija, variable o indefinida.

Fichero Secuencial Encadenado

- Un puntero es un campo adicional que almacena la dirección de otro registro en el fichero. Usando el puntero podemos acceder a ese registro.
- El uso de punteros permite recorrer los registros en un orden distinto al orden en que fueron registrados, estableciendo una secuencia lógica.
- Dado un registro, llamaremos "registro siguiente" a aquel al que se accede usando el puntero almacenado en el primero.
- Los registros quedarán ordenados mediante el orden que establece la secuencia de punteros.
- La ordenación generalmente está relacionada con el valor de algún otro campo.

Utiliza dos marcas especiales:

- **Puntero cabeceira:** Un puntero adicional que indica dónde comienza la secuencia.
- **Marca de final de ficheiro:** Indica dónde termina la secuencia.

Operaciones:

- **Recorrido del fichero:** Los registros se recorren secuencialmente utilizando los punteros, empezando por el puntero cabeceira y terminando en la marca de final de ficheiro.
- **Acceso a un registro:** Requiere crear un proceso en el que se recorre el fichero y se compara el valor del campo a localizar con el valor del mismo campo de los registros recorridos.
- **Consulta dun registro:** Requiere primero acceder al registro. Después la lectura es directa. Típicamente se harán consultas por el criterio de ordenación, por lo que bastará con seguir el puntero al registro siguiente.
- **Inserción dun registro:** La inserción física solo puede hacerse al final del fichero. Para insertar un registro entre los ya existentes, se insertará al final y se modificarán los punteros para mantener el orden.
- **Borrado dun registro:** Requiere primero acceder al registro. Después se modificará el valor del puntero del registro anterior para que apunte al

registro siguiente al que se quiere eliminar. El borrado, por tanto, es lógico.

- **Modificación dun registro:** No es posible la modificación. En su lugar se procederá a borrar el registro e insertarlo de nuevo, con el campo modificado.
- **Para mantener ordenado y compacto el fichero, es necesario realizar periódicamente operaciones de mantenimiento:** copiar los registros no borrados en otra posición de memoria y actualizar los punteros.

Fichero secuencial doblemente encadenado

Variación de los ficheros encadenados:

- Cada registro mantiene dos punteros:
 - Al registro siguiente.
 - Al registro anterior.
- Permite recorridos en ambos sentidos de la ordenación.
- Permite accesos más rápidos: se usarán los punteros en el sentido más conveniente.
- Facilita algunas operaciones, por ejemplo, el borrado.

Facilita el borrado porque cuando tienes dos punteros en cada registro que apuntan al registro anterior y al siguiente, puedes simplemente actualizar esos punteros para "saltar" el registro que deseas eliminar. Aquí hay una explicación más detallada:

- Las actualizaciones requieren actualizar ambos punteros.

Fichero de acceso directo o aleatorio

Organización con clave:

- Cada registro tiene un campo adicional clave, que permite localizarlo físicamente en el almacenamiento.
- Usando la clave se puede recuperar la posición física del registro en el almacenamiento y acceder a él directamente.
- Los registros pueden almacenarse en cualquier posición física del disco: acceso aleatorio.
- Suele usarse con sistemas de almacenamiento de memoria masiva de acceso directo, como discos magnéticos.

Tipos de de fichero directo

- Acceso directo: La clave contiene la dirección de memoria donde está el registro.
- Acceso directo indexado: Utiliza una tabla que relaciona índices con direcciones de memoria. La clave contiene un índice que se usa para recuperar en la tabla la dirección de memoria donde está el registro.
- Acceso directo calculado: Utiliza una función matemática. La clave contiene un valor que al serle aplicada dicha función permite obtener la dirección de memoria donde está el registro.

Otras características:

- Posicionamiento inmediato.

Los ficheros de acceso directo están diseñados para permitir el acceso inmediato a registros específicos utilizando una clave. Esto significa que puedes acceder directamente a los datos que necesitas sin tener que recorrer todos los registros en secuencia.

- Registros de longitud fija.

Para facilitar el acceso directo, se suelen utilizar registros de longitud fija. Esto significa que cada registro ocupa la misma cantidad de espacio en el almacenamiento. Dado que se utiliza una clave para ubicar registros, esta uniformidad en la longitud facilita la búsqueda y el acceso directo.

- Los registros se borran colocando un cero en la posición que ocupan.

La capacidad de borrar registros colocando un cero en la posición que ocupan se refiere a la eliminación lógica de registros. En lugar de eliminar físicamente un registro, se marca como borrado, lo que permite mantener la integridad de la estructura del fichero y facilita la recuperación si es necesario.

- Permiten el uso de algoritmos de compactación de huecos.

Cuando se eliminan registros, se crean huecos en el fichero. Los algoritmos de compactación de huecos permiten reorganizar el fichero para eliminar estos espacios vacíos y reducir el desperdicio de espacio de

almacenamiento. Esto ayuda a mantener el fichero eficiente a medida que se realizan operaciones de eliminación.

- En la creación del fichero se establece su tamaño máximo, es decir, el número máximo de registros que puede contener.

Establecer un tamaño máximo para el fichero significa que se reserva una cantidad específica de espacio en el almacenamiento para él. Esto puede ser útil para gestionar el uso de recursos y garantizar que el fichero no crezca indefinidamente.

- Se usan cuando el acceso a los datos de un registro siempre se hace con la misma clave y la velocidad de acceso a un registro es lo que más nos importa.
- Permiten la actualización de los registros en el mismo fichero, sin necesidad de copiarlo.

Los ficheros de acceso directo se utilizan cuando el acceso a los datos de un registro siempre se hace con la misma clave. Esto es especialmente eficiente cuando los datos se buscan y recuperan frecuentemente utilizando una clave específica.

- Permiten actualizar procesos en tiempo real.

Los ficheros de acceso directo permiten actualizar registros en el mismo fichero sin necesidad de copiarlo. Esto es útil cuando necesitas realizar actualizaciones en tiempo real en los datos, lo que puede ser crítico en aplicaciones que requieren procesamiento continuo.

Operaciones:

- **Recorrido del fichero:** Se recorrerá el índice secuencialmente y se accederá a cada registro directamente.
- **Acceso a un registro:** Acceso directo a partir de su clave utilizando el índice.
- **Consulta de un registro:** Requiere primero acceder al registro. Después la lectura es directa.
- **Inserción de un registro:** Si el índice está lleno, no se permite insertar más elementos. Si hay espacio, se inserta el elemento en la memoria y se almacena en el índice, en la posición correspondiente, su clave y su posición en memoria.

- **Borrado de un registro:** Se modifica el índice para que la posición del elemento con esa clave tome valor cero.
- **Modificación de un registro:** Requiere primero acceder al registro. Después se modifican los campos que corresponda.

Ficheros Indexados

Organización de registros en memoria:

- Los registros del fichero se guardan **secuencialmente** en memoria ordenados por uno de sus campos (**clave**).
- El fichero se divide lógicamente en **bloques de registros**. Todos los bloques tienen el mismo tamaño.
- Se utiliza un **índice en forma de tabla**, que relaciona la primera clave de cada bloque con la posición en memoria en la que se encuentra.
- Usando el índice se puede acceder de forma directa al **primer registro de cada bloque**.
- Una vez localizado el bloque, el acceso a los registros será **secuencial**.

Cada fichero consta de 3 áreas:

1. **Área primaria:** Contiene los registros ordenados secuencialmente por la clave y agrupados en bloques.
2. **Área de índices:** La crea el sistema en el momento de almacenar los datos. Cada registro contiene dos campos:
 - El valor de la clave del primer registro de un bloque.
 - La posición de memoria del primer registro del bloque.
3. **Área de overflow o de excedentes:** Contiene los registros que se añadieron al fichero después de su creación. Los registros no están ordenados por la clave y su clave puede intercalarse con las del área primaria.

Operaciones:

- **Recorrido del fichero:** No es fácil realizar un recorrido ordenado de los registros debido al área de overflow.
- **Acceso a un registro:** Puede requerir 2 o 3 pasos:
 - Utilizar el **índice** para acceder directamente al **primer registro de su bloque**.
 - Realizar un **recorrido secuencial** de los registros del bloque hasta encontrar el registro.
 - Si no se encuentra, realizar un **recorrido secuencial** de los registros del **área de overflow** hasta encontrarlo.
- **Consulta de un registro:** Requiere primero acceder al registro. Después la lectura es directa.
- **Inserción de un registro:** Se insertará al **final del área de overflow**. No se permiten inserciones en el área primaria después de la creación.

- **Borrado de un registro:** Requiere primero acceder al registro. Después el registro no se puede eliminar físicamente, sino que se marcará como borrado (**borrado lógico**).
- **Modificación de un registro:** No es posible la modificación. En su lugar se procederá a **borrar el registro e insertarlo de nuevo**, con el campo modificado.
- Para mantener ordenado y compacto el ficheiro, es necesario **reorganizar el archivo** cuando hay muchos **borrados** o si el **área de overflow es grande**.